eBird Best Practices

eBird Data Extraction and Processing in R

The Cornell Lab of Ornithology

eBird is a an online database of **bird observations** collected by **citizen scientists**, providing real-time data about the global **distribution and abundance** of bird species.

CHECKLIST S47044589

Mon 2 Jul 2018 1:11 PM

Edit Date and Effort

Winter Trail, Alaska, US (71.251, -156.491) North Slope County, Alaska, United States Edit Location

Matthew Strimas-Mackey Share

Andrew Spencer, Ian Davies, Nathan Pieplow

Traveling Complete

*Are you submitting a **complete checklist** of the birds you were able to identify?* **Yes**

Learn more

4 1 hr, 50 min 2.37 mi

eBird Mobile Tracks + Large map

Submitted from eBird Android, version 1.9

Edit comments

SUBMIT ANOTHER FOR...

Same location and date
Winter Trail, Alaska, US (71.251, -156.491), North Slope County, Alaska, US on Mon Jul 02, 2018

Same location
Winter Trail, Alaska, US (71.251, -156.491), North Slope County, Alaska, US

Same area and date
Another location near Winter Trail, Alaska, US (71.251, -156.491), North Slope County, Alaska, US on Mon Jul 02, 2018

Same area
Another location near Winter Trail, Alaska, US (71.251, -156.491), North Slope County, Alaska, US

Same date
Mon Jul 02, 2018

Different location and date

MEDIA POWERED BY MACAULAY LIBRARY

23
Species observed
+1 other taxa

2
Species with photos

150	Greater White-fronted Goose
1	Tundra Swan
15	Northern Pintail
1	Spectacled Eider BREEDING CODE: **NE Nest with Eggs (Confirmed)**
3	King Eider
25	Long-tailed Duck
120	American Golden-Plover Presumably failed breeders
12	Dunlin BREEDING CODE: **NE Nest with Eggs (Confirmed)**

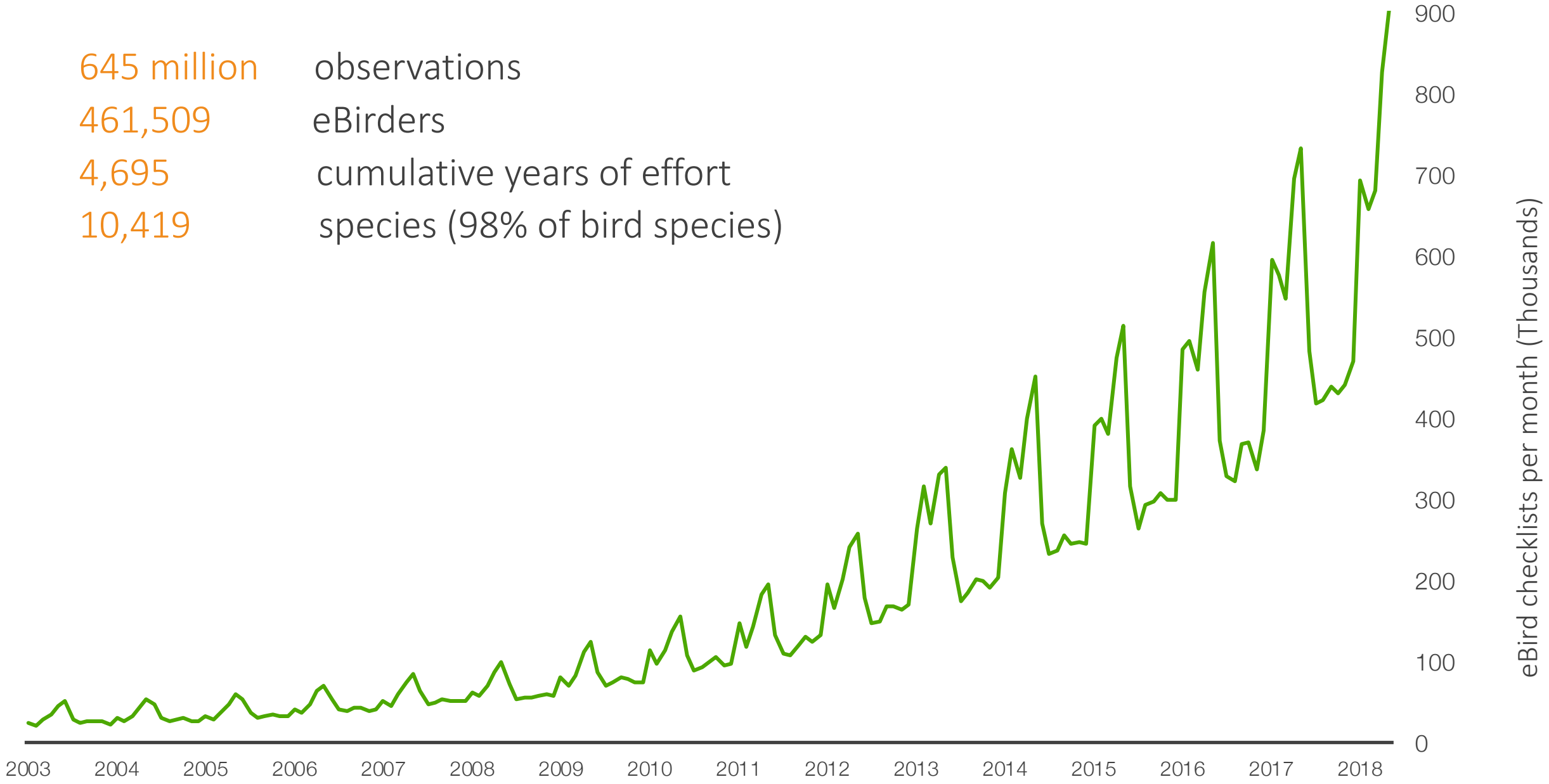
20	Pectoral Sandpiper BREEDING CODE: **NE Nest with Eggs (Confirmed)**

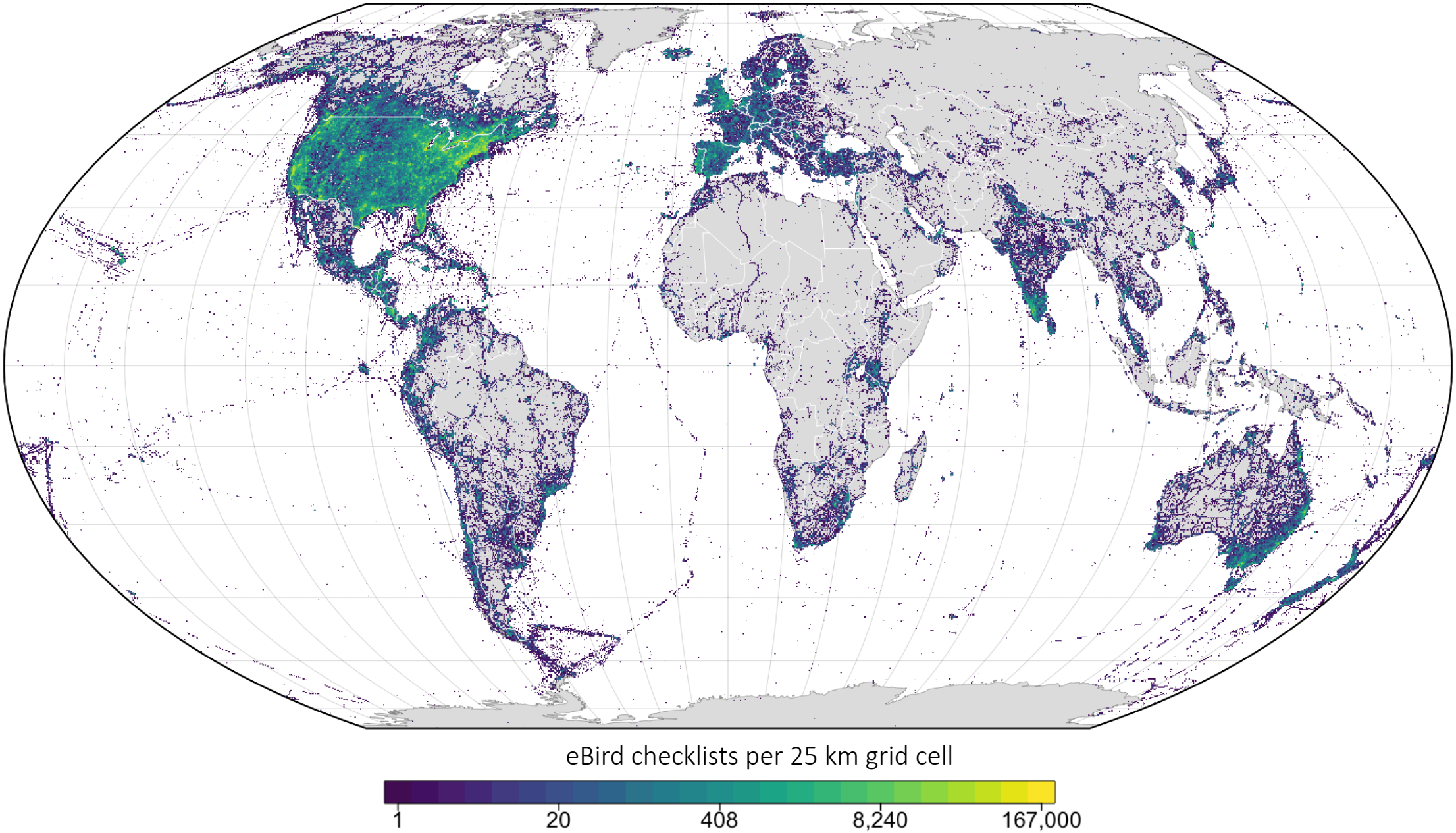
645 million observations

461,509 eBirders

4,695 cumulative years of effort

10,419 species (98% of bird species)




eBird checklists per 25 km grid cell
1
20
408
8,240
167,000

How do I access eBird data?

1. eBird website
2. eBird Status and Trends
3. Accessing raw eBird data

Species:

Date: Mar-May, Past 10 Years

Location: Enter place name or address...

Regina
Winnipeg
Seattle
WASHINGTON
Portland
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEBRASKA
MINNESOTA
Minneapolis
WISCONSIN
IOWA
Chicago
MICHIGAN
Detroit
Toronto
Ottawa
Montreal
Quebec City
NB
PRINCE EDWARD ISLAND
NOVA SCOTIA
MAINE
VERMONT
NEW HAMPSHIRE
MASSACHUSETTS
CT
RI
NEW YORK
New York
PENNSYLVANIA
Philadelphia
NJ
MARYLAND
DE
Washington
WEST VIRGINIA
VIRGINIA
OHIO
INDIANA
Indianapolis
ILLINOIS
Kansas City
MISSOURI
KENTUCKY
Nashville
TENNESSEE
NORTH CAROLINA
Charlotte
SOUTH CAROLINA
Atlanta
GEORGIA
ARKANSAS
MISSISSIPPI
ALABAMA
LOUISIANA
Dallas
Houston
New Orleans
Jacksonville
Orlando
Tampa
FLORIDA
Miami
Bermuda
The Bahamas
Havana
Gulf of Mexico
Mexico
TAMAULIPAS
DURANGO
CALIFORNIA SUR

1

Overview | Recent Visits | **Illustrated Checklist**

About

POWERED BY MACAULAY LIBRARY AND eBIRD

Jump to Species | Enter species name | All Media

234	26	2	208	232
Species observed	Species w/ photos	Species w/ audio	Species needing photos	Species needing audio

Bufflehead

Last seen: 30 Oct 2018 by David MacDougall

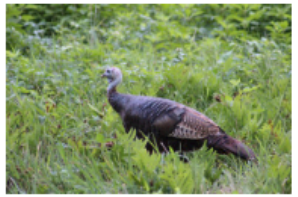
© Sally Chisholm

Audio needed
Contribute yours

More: 1

Wild Turkey

Last seen: 27 Jun 2018 by John Murphy



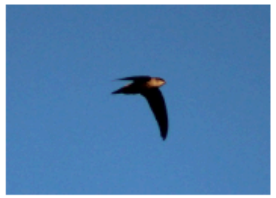
© Steven Lamonde

Audio needed
Contribute yours

More: 1

Chimney Swift

Last seen: 18 Jul 2018 by Christine Whitebread



© Steven Lamonde

Audio needed
Contribute yours

More: 1

Greater Yellowlegs

Last seen: 26 Oct 2018 by Ellen Freda

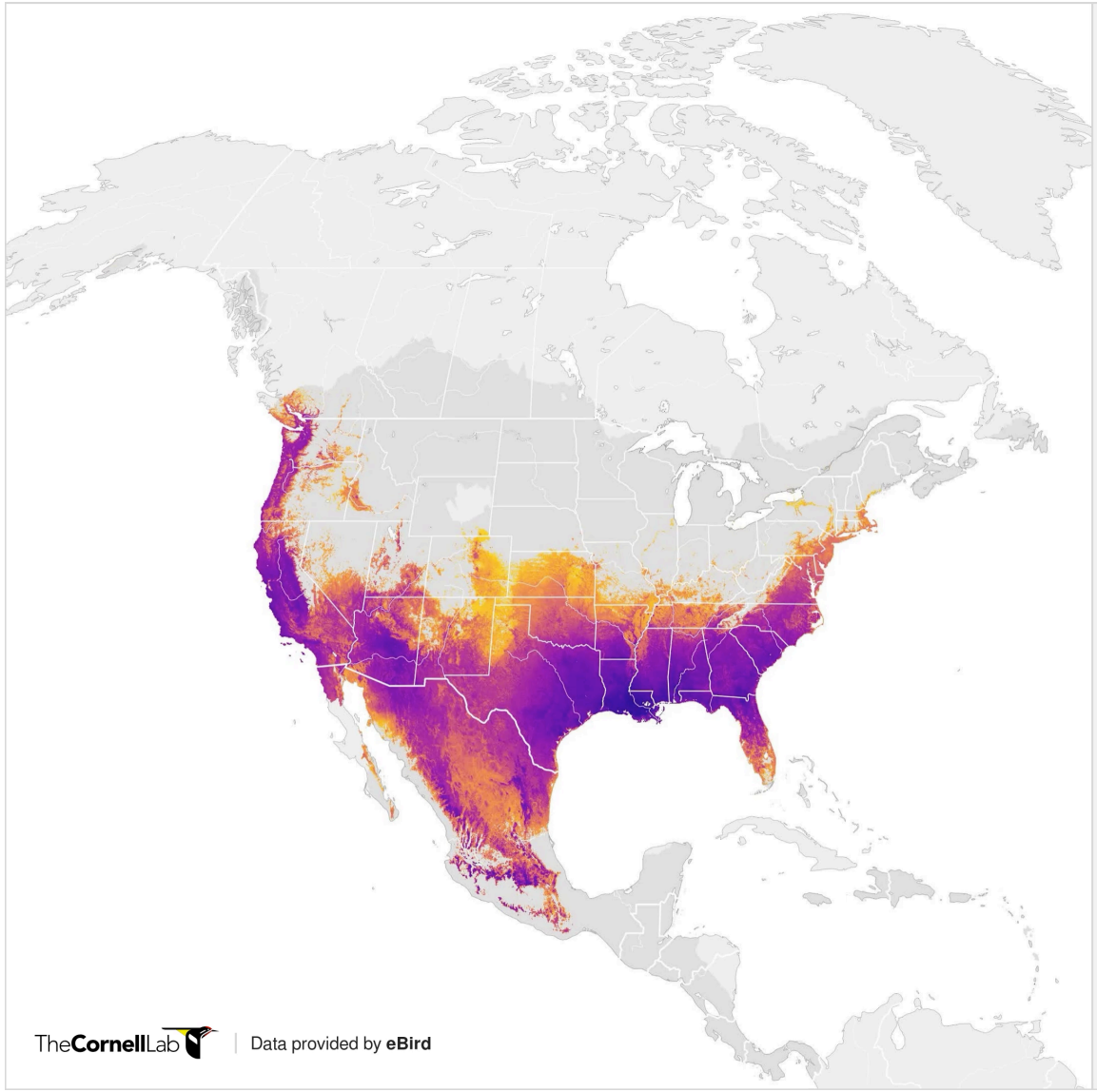
© Kenneth Scott

Audio needed
Contribute yours

More: 1

2

eBird
Status & Trends



Ruby-crowned Kinglet

Regulus calendula

Abundance

This map animates weekly estimated relative abundance, defined as the expected count on a one-hour, one kilometer eBird Traveling Count conducted at the ideal time of day for detection of that species in a region.

RELATIVE ABUNDANCE birds per km/hr

0.18 1.21 11.97

WEEK OF THE YEAR January 4

J F M A M J J A S O N D

Modeled area (0 abundance)

No prediction

eBird data from 2004-2016. Estimated for 2016.

Fink, D., T. Auer, A. Johnston, M. Strimas-Mackey, M. Iliff, and S. Kelling. eBird Status and Trends. Version: November 2018. https://ebird.org/science/status-and-trends. Cornell Lab of Ornithology, Ithaca, New York.

ebirdst: Access and Analyze eBird Status and Trends Data

Installation

Install `ebirdst` from CRAN with:

```
install.packages("ebirdst")
```

Alternatively, you can install the development version from GitHub with:

```
# install.packages("remotes")
remotes::install_github("CornellLabofOrnithology/ebirdst")
```

Vignettes

For a full introduction and advanced usage, please see the package website. An introductory vignette details the data access and structure of the results. An intro mapping vignette expands upon the quick start readme and shows the basic mapping moves. The advanced mapping vignette shows how to reproduce the seasonal maps and statistics on the eBird Status and Trends website. Finally, the non-raster data vignette details how to access additional information from the model results about predictor importance and directionality, as well as predictive performance metrics.

Quick Start

This quick start guide shows how to download example data and plot abundance values similar to how they are plotted for the eBird Status and Trends Abundance animations. **Important note: after downloading the results, do not change the file structure.** All functionality in this package relies on the structure inherent in the delivered results. Changing the folder and file structure will cause errors with this package. If you use this package to analyze the results, you do not ever need to interact with the files directly, outside of R.

```
library(ebirdst)
library(viridis)
library(raster)
library(sf)
library(rnaturalearth)
```

3

Home > Science

Using eBird for science

Interested in using eBird data scientifically? Here are our key tips and tricks and the next opportunities to learn from the eBird analysis team in person.

Upcoming workshops

Members of the eBird analysis team are leading workshops at the American Ornithological Society Conference in Alaska in June 2019. See all workshops.

Best Practices for eBird Data I: Accessing and preparing eBird data for analysis in R

Tuesday 25 June 8am – 12 noon

Best Practices for eBird Data II: Modeling distribution and abundance using eBird data

Tuesday 25 June 1pm – 5pm

Working with eBird Status and Trends data products in R

Friday 28 June 12 noon – 2pm

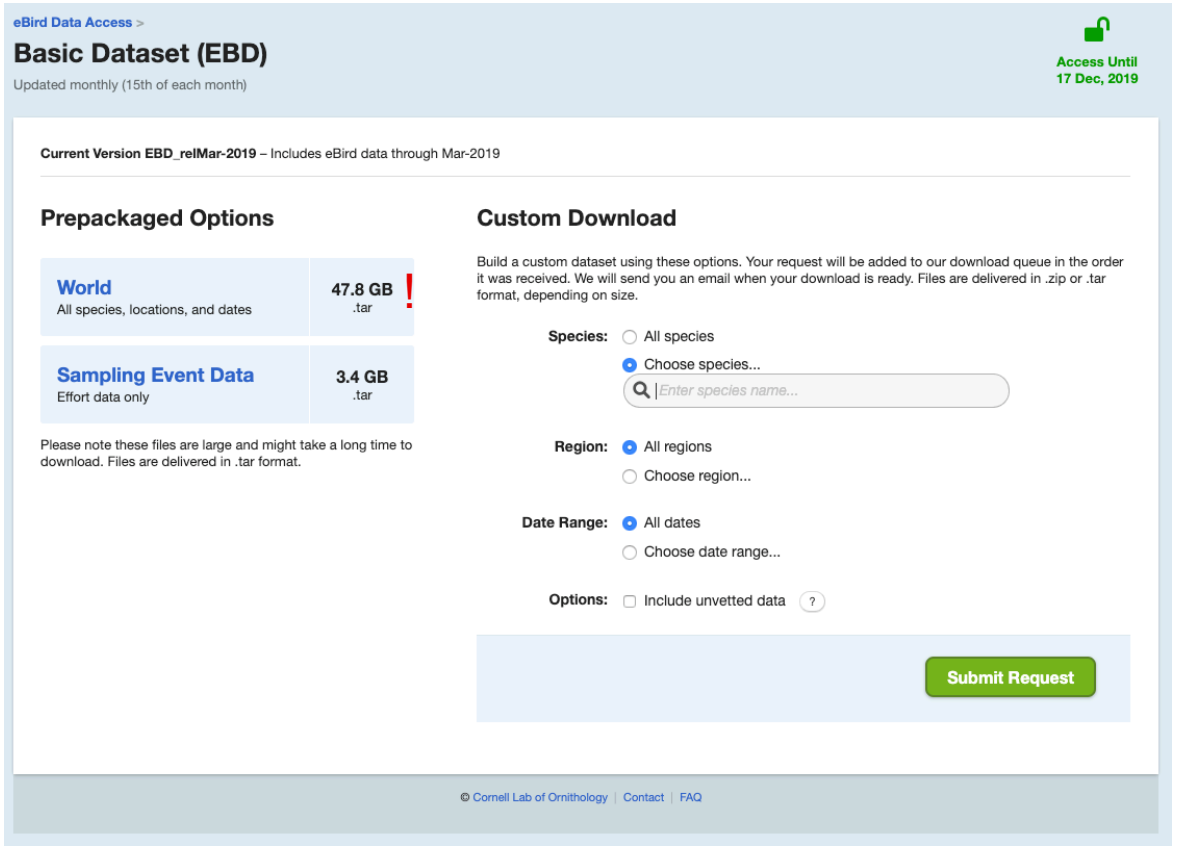
Accessing raw eBird data

Download raw data here.

Once you have the data in hand, the R package *auk* can be used to extract and prepare eBird data for analysis. Learn about *auk* here.

Analysing raw eBird data

eBird data bring a number of challenges for analysis and it is important to consider how the data are generated when using them for analysis. We outline best practices for analysing eBird data in this paper. Don't forget to look at the appendices! We also provide R code for producing species distributions with eBird here: http://strimas.com/ebird-best-practices/. Many of the principles outlined here apply to other analyses with eBird and other citizen science datasets.



Name	Date Modified	Size	Kind
ebd_relFeb-2018.txt	Today, 6:38 PM	183.16 GB	Plain Text
eBird_Basic_Dataset_Metadata_v1.9.pdf	3/13/18, 10:47 AM	165 KB	PDF
IBACodes.txt	11/1/12, 6:13 PM	127 KB	Plain Text
USFWSCodes.txt	3/17/17, 10:33 AM	40 KB	Plain Text
terms_of_use.txt	3/12/15, 4:27 PM	7 KB	Plain Text
BCRCodes.txt	6/9/14, 10:46 AM	2 KB	Plain Text
recommended_citation.txt	12/13/17, 12:08 PM	103 bytes	Plain Text

We need to make this

FILE

smaller

ebd_relFeb-2018 — -bash — 116×28

```
[ag-clo-mes335:ebd_relFeb-2018 mes335$ cd /Users/mes335/data/ebird/ebd_relFeb-2018
[ag-clo-mes335:ebd_relFeb-2018 mes335$ ls
awk                                     ebd_relFeb-2018_subset.txt
ebd_relFeb-2018_SG.txt                  ebd_sampling_relFeb-2018_SG.txt
ebd_relFeb-2018_clean.txt               ebd_sampling_relFeb-2018_clean.txt
ebd_relFeb-2018_rollup.txt              ebird_country-state.txt
[ag-clo-mes335:ebd_relFeb-2018 mes335$ head -20 ebd_relFeb-2018_clean.txt | awk -F $'\t' '{print $25,$26,$27,$5,$9'} ]
LATITUDE LONGITUDE OBSERVATION DATE COMMON NAME OBSERVATION COUNT
35.1085767 -120.6265268 2008-11-05 Belted Kingfisher 1
35.1085767 -120.6265268 2008-11-05 Wilson's Snipe 1
35.1085767 -120.6265268 2008-11-05 Ruby-crowned Kinglet 15
35.1085767 -120.6265268 2008-11-05 California Gull 35
35.1085767 -120.6265268 2008-11-05 Mallard 40
35.1085767 -120.6265268 2008-11-05 Bewick's Wren 1
8.52111 -83.39991 2004-06-18 Rufous Piha X
8.52111 -83.39991 2004-06-18 Thick-billed Euphonia X
46.3024735 -91.0284233 2008-08-16 Chestnut-sided Warbler 4
46.3024735 -91.0284233 2008-08-16 American Robin 9
33.7968382 -117.7532673 1971-01-02 Spotted Towhee 4
33.7968382 -117.7532673 1971-01-02 Red-tailed Hawk 5
33.7968382 -117.7532673 1971-01-02 Anna's Hummingbird 11
33.7968382 -117.7532673 1971-01-02 Bewick's Wren 5
37.1434192 -114.0232372 2008-10-25 Spotted Towhee 6
43.1812545 -88.0415756 2008-08-06 Mallard 5
43.1812545 -88.0415756 2008-08-06 Red-winged Blackbird 3
41.059032 -112.238045 2006-08-13 Black-necked Stilt 1
41.059032 -112.238045 2006-08-13 California Gull 1
ag-clo-mes335:ebd_relFeb-2018 mes335$
```

Terminal 😐

AWK 😟

awk-example.txt

```
BEGIN {
  FS = OFS = "  "
    split("Cyanocitta cristata  Perisoreus canadensis", speciesValues, "  ")
    for (i in speciesValues) species[speciesValues[i]] = 1
    split("CA US", countryValues, " ")
    for (i in countryValues) countries[countryValues[i]] = 1
}
{
  keep = 1
  if (keep == 1 && ($6 in species)) {
    keep = 1
  } else {
    keep = 0
  }
  if (keep == 1 && ($14 in countries)) {
    keep = 1
  } else {
    keep = 0
  }
  if (keep == 1 && ($26 >= -100 && $26 <= -80 && $25 >= 37 && $25 <= 52)) {
    keep = 1
  } else {
    keep = 0
  }
  if (keep == 1 && ($27 >= "2012-01-01" && $27 <= "2012-12-31")) {
    keep = 1
  } else {
    keep = 0
  }
  if (keep == 1 && ($28 >= "06:00" && $28 <= "09:00")) {
    keep = 1
```

auk 0.3.3 Vignette Function Reference

auk: eBird Data Extraction and Processing in R

License GPL v3 | Mac OSX & Linux passing | Windows passing | coverage 79%
Peer Reviewed | CRAN 0.3.2 | downloads 13K

Overview

eBird is an online tool for recording bird observations. Since its inception, over 600 million records of bird sightings (i.e. combinations of location, date, time, and bird species) have been collected, making eBird one of the largest citizen science projects in history and an extremely valuable resource for bird research and conservation. The full eBird database is packa as a text file and available for download as the eBird Basic Dataset (EBD). Due to the large size of this dataset, it must be filtere smaller subset of desired observations before reading into R. This filtering is most efficiently done using AWK, a Unix utility a programming language for processing column formatted text data. This package acts as a front end for AWK, allowing user eBird data before import into R.

Installation

```
# cran release
install.packages("auk")

# or install the development version from github
# install.packages("remotes")
remotes::install_github("CornellLabofOrnithology/auk")
```

`auk` requires the Unix utility AWK, which is available on most Linux and Mac OS X machines. Windows users will first need to install Cygwin before using this package. Note that **Cygwin must be installed in the default location** (`C:/cygwin/bin/gawk.exe` or `C:/cygwin64/bin/gawk.exe`) in order for `auk` to work.

Vignette

Full details on using `auk` to produce both presence-only and presence-absence data are outlined in the vignette, which can be accessed with `vignette("auk")`.

eBird Best Practices

Best Practices for Using eBird Data

Matthew Strimas-Mackey, Alison Johnston, Wesley M. Hochachka, Viviana Ruiz-Gutierrez, Orin J. Robinson, Eliot T. Miller, Tom Auer, Steve Kelling, Daniel Fink

2019-04-18

Welcome

This book is currently under active development. Please explore and give us feedback, but be aware that some sections are incomplete and others may still contain bugs.

Best Practices for Using eBird Data is a supplement to *Best practices for making reliable inferences from citizen science data: case study using eBird to estimate species distributions* (Johnston et al. 2019). This paper describes the challenges associated with making inferences from biological citizen science data and proposes a set of best practices for making reliable estimates of species distributions from these data. Throughout, the paper uses eBird, the world's largest biological citizen science project, as a case study to illustrate the good practices. This book acts as a supplement to the paper, showing readers how to implement these best practices within R using real data from eBird. After completing this book, readers should be able to extract data from the eBird database suitable for their own studies, process these data to prepare them for robust analyses, collect environmental covariates for modeling, and fit and assess models estimating encounter rate, occupancy, and relative abundance. Readers should be comfortable with the R programming language, and read the Prerequisites and Setup sections of the introduction, before diving into this book.

A preprint of the paper associated with this book is currently available on bioRxiv.

auk workflow

1. data access
2. filter
3. import
4. pre-process
5. presence-absence data

Applications

1. mapping and summarizing
2. occupancy model preparation
3. land cover covariates